Am. J. Hum. Genet. 44:388–396, 1989

Abundant Class of Human DNA Polymorphisms Which Can Be Typed Using the Polymerase Chain Reaction

James L. Weber and Paula E. May

Marshfield Medical Research Foundation, Marshfield, WI

Summary

Interspersed DNA elements of the form $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ constitute one of the most abundant human repetitive DNA families. We report that specific human $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ blocks are polymorphic in length among individuals and therefore represent a vast new pool of potential genetic markers. Comparison of sequences from the literature for $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ blocks cloned two or more times revealed length polymorphisms in seven of eight cases. Variations in the lengths of 10 $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ blocks were directly demonstrated by amplifying the DNA within and immediately flanking the repeat blocks by using the polymerase chain reaction and then resolving the amplified DNA on polyacrylamide DNA sequencing gels. Use of the polymerase chain reaction to detect DNA polymorphisms offers improved sensitivity and speed compared with standard blotting and hybridization.

Introduction

A subclass of eukaryote tandemly repeated DNA contains very short simple sequence repeats such as $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$, hereafter designated $(CA)_n$ (Miesfeld et al. 1981; Hamada and Kakunaga 1982; Sun et al. 1984; Tautz and Renz 1984; Gross and Garrard 1986; Pardue et al. 1987; Braaten et al. 1988). In the human genome there are 50,000–100,000 interspersed $(CA)_n$ blocks, with the range of n being roughly 15–30 (Hamada and Kakunaga 1982; Hamada et al. 1982, 1984*a*; Sun et al. 1984; Tautz and Renz 1984; Gross and Garrard 1986; Braaten et al. 1988). Uniform spacing of the $(CA)_n$ blocks throughout the genome would place them every 30–60 kb. Function for the $(CA)_n$ blocks is unknown, but it has been proposed that they serve as hot spots for recombination (Slightom et al. 1980) or participate in gene regulation (Hamada et al. 1984*b*). $(CA)_n$ blocks can adopt the Z DNA conformation within negatively supercoiled plasmids (Nordheim and Rich 1983; Hamada et al. 1984*a*), but current information indicates they are not in the Z DNA conformation within cells (Tautz and Renz 1984; Gross et al. 1985).

Several examples are known for which the number of repeats within a block of tandemly repeated DNA varies among individuals within a species. These examples include primate alphoid satellite DNA (Tyler-Smith and Brown 1987; Waye et al. 1987), human color-vision genes (Motulsky 1988), tandem repeats within malaria parasite antigen genes (Kemp et al. 1987; Weber 1988), and minisatellites (Jeffreys et al. 1985; Nakamura et al. 1987). By analogy with these examples we hypothesized that blocks of human $(CA)_n$ repeats exhibit length polymorphisms.

Material and Methods

Data-base searches and sequence analyses were performed using software made available by IntelliGenetics Inc. (Mountain View, CA) through the national BIONET computing network.

Genomic DNA was isolated from nucleated blood cells as described elsewhere (Aldridge et al. 1984). Standard polymerase chain reactions (Saiki et al. 1985, 1988; Mullis and Faloona 1987) were carried out in a 25-μl vol containing 10–20 ng of genomic DNA template, 100 ng each oligodeoxynucleotide primer, 200 μM each

Received September 26, 1988; revision received November 14, 1988.

Address for correspondence and reprints: James L. Weber, Marshfield Medical Research Foundation, 510 North Saint Joseph Avenue, Marshfield, WI 54449.

0002-9297/89/4403-0012$02.00

dGTP, dCTP, and dTTP, 2.5 μM dATP, 1–2 μCi α^{32}P-dATP at 800 Ci/mmol or α^{35}S-dATP at 500 Ci/mmol, 50 mM KCl, 10 mM Tris (pH 8.3), 1.5 mM $MgCl_2$, 0.01% gelatin, and about 0.75 unit of *Taq* polymerase (Perkin Elmer Cetus, Norwalk, CT). Samples were overlaid with mineral oil and were processed through 25 temperature cycles consisting of 1 min at 94°C (denaturation), 2 min at 55°C (annealing), and 2.5 min at 72°C (elongation). The last elongation step was lengthened to 10 min. Aliquots of the amplified DNA were mixed with 2 vol formamide sample buffer and were electrophoresed on standard denaturing polyacrylamide DNA sequencing gels (Biggin et al. 1983). Gels were then fixed, dried, and processed for autoradiography. Exposure times were about 2 d. Gel size standards were dideoxy sequencing ladders produced by using m13, mp10 DNA as template. Oligodeoxynucleotide primers were synthesized on a Cyclone DNA synthesizer (Biosearch, Inc., San Rafael, CA).

Results shown in figure 1 were obtained using conditions slightly different than the standard conditions. Templates were 100–200 ng of genomic DNA, annealing steps were 2.5 min at 37°C, elongation steps were 3.5 min at 72°C and α^{35}S-dATP was added after the 18th cycle rather than at the beginning of the reactions. The plasmid DNA sample was amplified starting with 50 pg of total plasmid DNA as template.

The insulin-like growth factor I (IGF1) CA strand oligodeoxynucleotide primer was end-labeled for 1 h at 37°C in a 50-μl reaction containing 72 pmol (480 ng) primer, 33 pmol γ^{32}P-ATP at 3,000 Ci/mmol, 10 mM $MgCl_2$, 5 mM DTT, 50 mM Tris (pH 7.6); and 50 units T4 polynucleotide kinase. Polymerase chain reactions were carried out in 50 μl volumes with 100 ng of end-labeled primer and 300 ng of each unlabeled primer.

Genomic DNA or total DNA from a chromosome 19–specific large insert phage library (LL19NL01; gift of P. de Jong and M. Van Dilla) was digested to completion with *Sau*3A and *Taq*I. DNA fragments ranging in size from about 150 to 350 bp were purified by preparative agarose gel electrophoresis (Weber et al. 1988) and were ligated into the mp18, ml3 vector. Nitrocellulose plaque lifts (Benton and Davis 1977) prepared from the resulting clones were screened by hybridization to synthetic poly(dC-dA) · poly(dG-dT) which had been nick-translated, using both α^{32}P-dATP and α^{32}P-dTTP, to a specific activity of about 5×10^7 cpm/μg. Hybridizations were carried out in 6 × SSC (pH 7.0), 2.5 mM EDTA, 5.0% (v/v) O'Darby Irish Cream Liqueur at 65°C overnight. Filters were washed at 65°C in 2 × SSC, 25 mM $NaPO_4$, 0.10% SDS, 5.0 mM EDTA, 1.5 mM $Na_4P_2O_7$ (pH 7.0), and then in 1 × SSC, 0.10% SDS, 5.0 mM EDTA (pH 7.0). Single-stranded DNA was isolated from the positive clones and was sequenced as described elsewhere (Biggin et al. 1983).

Estimates of PIC (Botstein et al. 1980) and heterozygosity were obtained either by typing DNA from 41–45 unrelated Caucasians for markers Mfd1–Mfd4 or by typing DNA from 75–78 parents of the 40 Centre d'Étude du Polymorphisme Humain (CEPH) reference families for markers Mfd5–Mfd10. The CEPH families are from the United States, France, and Venezuela.

Results

All human DNA sequences within GenBank, version 54, were screened for the presence of $(CA)_n$ [or $(GT)^\circ_n$] sequences with $n \geqslant 6$. A total of 59 $(CA)_n$ blocks were found. Two of the total, containing 6 and 8 repeats, were located within protein-coding regions; all of the other blocks were found within introns or between genes. About half of the repeat blocks contained one or more imperfections, such as an extra base or a base substitution, in the run of repeats, and about 40% of the repeat blocks were flanked by runs of other dinucleotides, such as CT or TA. Half of the blocks contained 13 or more CA repeats, and 7 of the 59 blocks contained 20 or more repeats. The longest block, found within an intron of the rhodopsin gene, contained 30 repeats.

Several of the $(CA)_n$ blocks found in GenBank have been sequenced two or more times from independent clones. The eight examples of this type found to date are listed in table 1. Surprisingly, all but one example showed variation in the number of repeats per block. Several investigators have noted these differences for individual cases (Slightom et al. 1980; Shen and Rutter 1984; Das et al. 1987).

To confirm and extend the results from the literature sequences, 10 $(CA)_n$ blocks were chosen for direct experimental examination (table 2). Sequences for 5 of the 10 test $(CA)_n$ blocks (Mfd1–Mfd5) were taken from GenBank, while the other five (Mfd6–Mfd10) were determined using human genomic DNA clones selected by hybridization to nick-translated poly(dC-dA) · poly(dG-dT). DNA within and immediately flanking the test $(CA)_n$ blocks was amplified and labeled using the polymerase chain reaction and then was electrophoresed on standard denaturing polyacrylamide DNA sequencing gels.

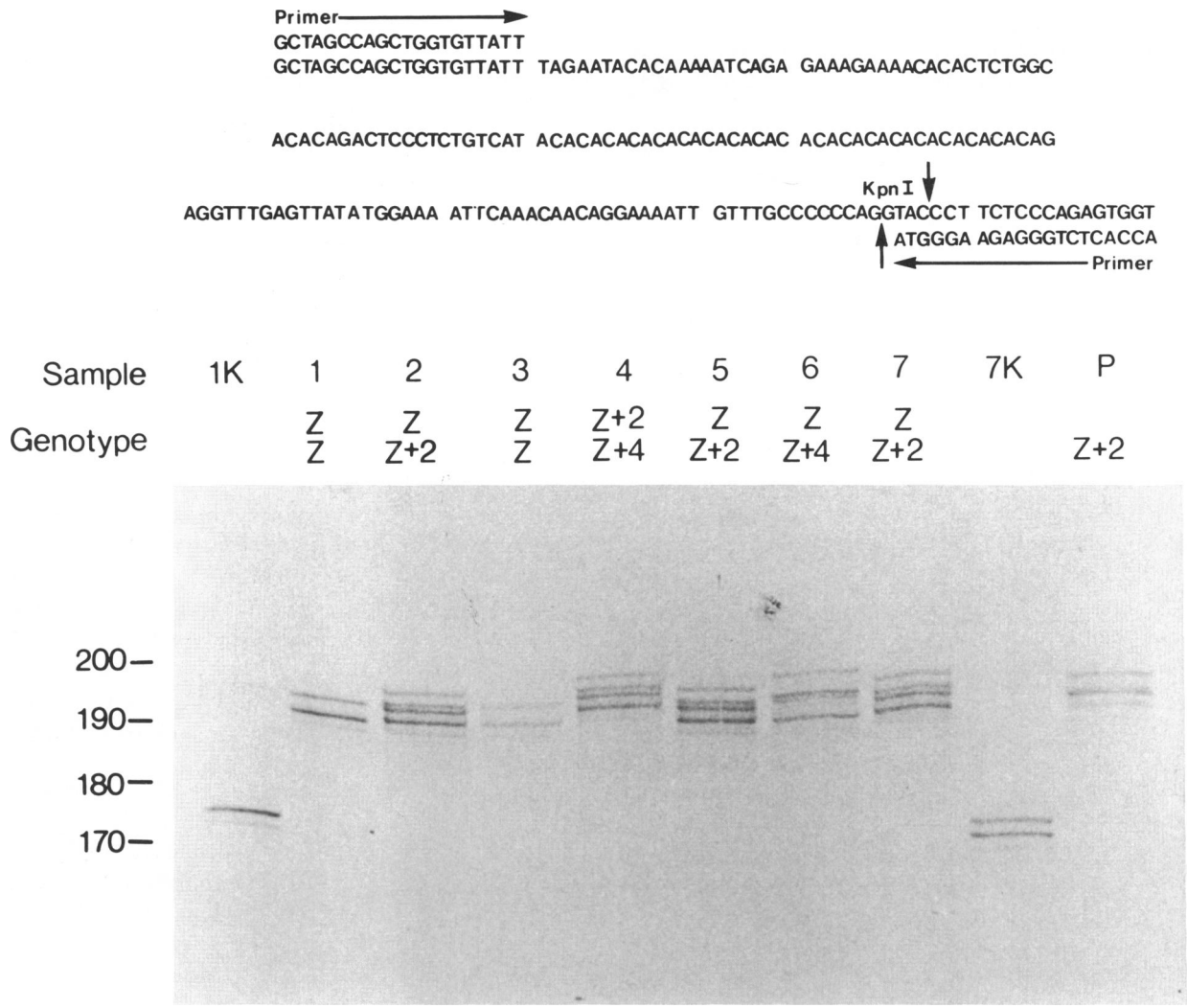

Figure 1 Genotypes for the IGF1 $(CA)_n$ block in seven unrelated individuals (1–7). Z = the most frequent allele; Z + 2 = the allele that is 2 bp larger than the most frequent allele; Z − 2 = the allele that is 2 bp smaller, etc. K = *Kpn*I digestion of amplified samples 1 and 7. *Kpn*I digestions reduce the number of bands to half the original number because the CA strand, which normally migrates with an apparent size of about 4 bases less than that of the GT strand, is, after *Kpn*I digestion, 4 bases longer than the GT strand, resulting in comigration of the two strands. P = DNA amplified from the plasmid DNA sample containing the IGF1 $(CA)_n$ block. At the top of the figure is shown the sequence of the amplified DNA segment along with the amplification primers and the *Kpn*I site. Sizes of the DNA fragments (in bases) are indicated on the left.

Amplified DNA from the IGF1 $(CA)_n$ block from seven unrelated individuals is shown in figure 1 (lanes 1–7). Because the CA and GT strands of the amplified DNA fragment migrate with different mobilities under the denaturing electrophoresis conditions (see below), homozygotes yield two bands and heterozygotes yield four bands. The band corresponding to the faster-moving CA strand is more intense on the autoradiographs than is the band for the slower GT strand because the adenine content of the CA strand is higher and because labeling is with α^{32}P-dATP. Two of the seven individuals shown in figure 1 (lanes 1 and 3) were homozygous for the predominant allele (Z) of the IGF1 $(CA)_n$ block; the remainder were heterozygotes of various types.

The amplified DNA fragments shown in figure 1 were the size that, on the basis of the DNA sequence, was expected for the IGF1 $(CA)_n$ block, and they matched in size the DNA amplified from a plasmid containing the IGF1 $(CA)_n$ block (Rotwein et al. 1986). The am-

Table 1

Literature Variations in Lengths of Human $(CA)_n$ Blocks

Gene	No. of Repeats	Repeat Sequences
Apolipoprotein AII	19	$(GT)_{19}$
	16, 16, 16	$(GT)_{16}$
Apolipoprotein CII	30	$(TG)_{23}(AG)_7$
	29	$(TG)_{22}(AG)_7$
	17	$(TG)_9(AG)_8$
c-sis	12	$(CA)_2TA(CA)_7TACA$
	12	$(CA)_2TA(CA)_7TACA$
Globin $^G\gamma$	23	$(TG)_{11}(CG)_3TT(TG)_6$
	32	$(TG)_{18}(CG)_3AC(TG)_3TT(TG)_6$
Globin $^A\gamma$	13	$(TG)_{13}$
	22	$(TG)_9(CG)_5(TG)_8$
Globin δ-β	19	$(TG)_{17}T(TG)_2$
Intergenic region	18	$(TG)_{16}T(TG)_2$
Somatostatin	25	$(TG)_{13}TT(TG)_2CG(TG)_8$
	28	$(TG)_{13}TT(TG)_2CG(TG)_{11}$
T-cell receptor	20	$(GT)_3TT(GT)_{16}$
Delta chain	22	$(GT)_3TT(GT)_{18}$

NOTE.—The number of repeated dinucleotides along with the sequences of the repeated regions are listed on separate lines for different alleles. Three of the four apolipoprotein AII alleles contained 16 repeats.

Table 2

$(CA)_n$ Block Markers

Marker[a]	Locus	Chromosome Location	Length of Amplified DNA[b] (bp)	Repeat Sequence[c]	PCR Primers[d]	No. of Alleles	% Hetero-zygotes	PIC
Mfd1	IGF1	12q22-q24.1	192	$CATA(CA)_{19}$	GCTAGCCAGCTGGTGTTATT ACCACTCTGGGAGAAGGGTA	5	54	.53
Mfd2	RHO	3q21-qter	120	$(CA)_{13}A(CA)_{17}$	CATTAGGATGCATTCTTCTG GTCAGGATTGAACTGGGAAC	4	34	.31
Mfd3	APOA2	1q21-q23	137	$(AC)_{16}$	GGTCTGGAAGTACTGAGAAA GATTCACTGCTGTGGACCCA	6	74	.65
Mfd4	SST	3q28	169	$(AC)_{12}GCACAA(AC)_{13}$	GCTCAAATGTTTCTGCAACC CTTTGTAGCTCGTGATGTGA	6	51	.46
Mfd5	APOC2	19q12-q13.2	151	$(CT)_7(CA)_{23}$	CATAGCGAGACTCCATCTCC GGGAGAGGGCAAAGATCGAT	11	80	.79
Mfd6	...	...	192	$(CA)_5AA(CA)_{13}$	TCCTACCTTAATTTCTGCCT GCAGGTTGTTTAATTTCGGC	7	49	.50
Mfd7	...	...	213	$(CA)_{20}TA(CA)_2$	GTTAGCATAATGCCCTCAAG CGATGGAGTTTATGTTGAGA	6	54	.51
Mfd8	...	...	185	$(AC)_{20}$	CGAAAGTTCAGAGATTTGCA ACATTAGGATTAGCTGTGGA	8	58	.58
Mfd9	...	19[e]	100	$(CA)_{17}$	GATGTCTCCTTGGTAAGTTA AATACCTAGGAAGGGGAGGG	9	72	.69
Mfd10	...	19	138	$(AC)_{14}$	CATGCCTGGCCTTACTTGC AGTTTGAGACCAGCCTGCG	6	39	.42

[a] Mfd = Marshfield. The marker name substitutes for the probe name in RFLP markers and denotes a specific pair of amplification primers for each locus.

[b] Sizes of amplified DNA fragments corresponded to the predominant allele for each marker. Estimated error is 2 bp.

[c] Sequences for Mfd1–Mfd5 were taken from GenBank. Sequences for Mfd6–Mfd10 were determined in the laboratory.

[d] The CA strand primer is listed first for each pair.

[e] Markers 9 and 10 were developed using clones selected from a chromosome 19–specific phage library.

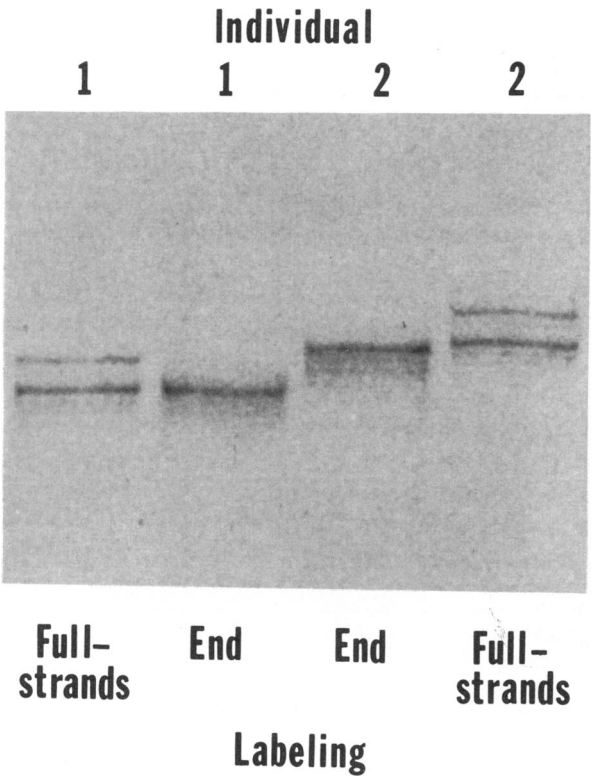

Figure 2 End labeling of amplified DNA. Amplified DNA from the IGF1 $(CA)_n$ block from two individuals was labeled either by incorporating α^{32}P-dATP throughout the length of both strands (Full-strands) or by using a ^{32}P–end labeled CA strand primer (End). Individual 1 is homozygous for the Z allele; individual 2 is homozygous for the Z + 4 allele.

plified DNA also hybridized to nick-translated poly(dC-dA) · poly(dG-dT) (not shown) and was cleaved by a restriction enzyme, *Kpn*I, at the expected position (fig. 1, lanes 1K and 7K). Similar diagnostic restriction-enzyme digestions were successfully performed on the amplified DNA from three other test $(CA)_n$ blocks: *Hpa*II for the Rhodopsin (RHO) block, *Pvu*II for the apolipoprotein AII (APOA2) block, and *Acc*I for the somatostatin (SST) block.

Rather than the amplified DNA being labeled throughout the interiors of both strands, one or both of the polymerase chain reaction primers can be end-labeled using polynucleotide kinase. Comparison of the two labeling schemes, for the IGF1 $(CA)_n$ block, is shown in figure 2. End-labeling the CA strand primer resulted in the appearance of only the higher-mobility band in two homozygous individuals with different genotypes.

All 10 test $(CA)_n$ blocks exhibited polymorphism in the sizes of the amplified fragments. Alleles always differed in size by multiples of 2 bases (from CA strand to CA strand bands), a result consistent with the concept that the number of tandem dinucleotide repeats is the variable factor. Allele frequencies for the 10 test markers are shown in table 3. For most of the test markers, major alleles were clustered in size, within about 6 bp, on either side of the predominant allele. Amplified fragments must therefore be small enough so that alleles differing in size by as little as 2 bases can easily be resolved on the polyacrylamide gels. Note also that the size differences between the largest and smallest alleles were $\leqslant$20 bp for most of the markers and that therefore several markers can be analyzed simultaneously on the same gel lane.

To determine whether the $(CA)_n$ block markers exhibited Mendelian codominant inheritance, DNA samples from four 3-generation families were amplified using primers for the first four markers. An example, for the IGF1, SST, and APOA2 markers, is shown in figure 3. All markers showed Mendelian behavior for all families; no new mutations were found in a total of 176 independent meioses. Jeffreys et al. (1988) recently reported that new mutations in hypervariable minisatellites were detected in families only for markers with $\geqslant$97% heterozygosity. If this result is applicable to the $(CA)_n$ markers, then new mutations in these markers are unlikely to be a general problem.

Additional bands, less intense than the major pair of bands for each allele and smaller in size than the major bands, were usually seen for the amplified DNA fragments. These bands are particularly apparent in figure 3. The additional bands were present when cloned DNA versus genomic DNA was used as template (fig. 1, lane P)—and even when such small amounts of heterozygote genomic DNA were used as template that only one of the two alleles was amplified. Also, DNA amplified from 63 lymphocyte clones (gift of J. Nicklas) (Nicklas et al. 1987) produced from two individuals showed no variation in genotype for all clones from a single donor. Taken together, these results strongly indicate that the additional bands are generated during the amplification reaction and that they are not, at least primarily, a reflection of somatic mosaicism.

The informativeness of the $(CA)_n$ block markers was generally superior to that of standard unique-sequence probe polymorphisms (Gilliam et al. 1987; Overhauser et al. 1987; Schumm et al. 1988) and was as good as that of many minisatellite polymorphisms (Nakamura

Table 3

Allele Frequencies for $(CA)_n$ Block Markers

	Marker[b]									
Allele[a]	1	2	3	4	5	6	7	8	9	10
Z + 14	–	–	–	–	1%	–	–	–	–	–
Z + 10	–	–	–	–	–	3%	–	1%	–	–
Z + 8	–	–	3%	–	3%	11%	–	3%	–	–
Z + 6	–	–	35%	2%	4%	1%	–	7%	3%	1%
Z + 4	18%	6%	–	6%	15%	1%	6%	2%	1%	–
Z + 2	16%	11%	1%	13%	15%	3%	4%	27%	12%	5%
Z	60%	81%	39%	70%	34%	64%	66%	54%	38%	73%
Z – 2	5%	2%	–	1%	3%	18%	15%	4%	6%	–
Z – 4	–	–	8%	–	2%	–	7%	1%	4%	10%
Z – 6	–	–	14%	8%	1%	–	–	–	4%	9%
Z – 8	–	–	–	–	–	–	2%	–	31%	1%
Z – 12	–	–	–	–	–	–	–	–	1%	–
Z – 14	–	–	–	–	11%	–	–	–	–	–
Z – 16	1%	–	–	–	–	–	–	–	–	–
Z – 22	–	–	–	–	12%	–	–	–	–	–
Spread[c]	20 bp	6 bp	14 bp	12 bp	36 bp	12 bp	12 bp	14 bp	18 bp	14 bp

[a] Z = the predominant allele for each marker as described in the legend to fig. 1.
[b] A dash (–) denotes that this allele was not found.
[c] Size difference between amplified DNA fragments corresponding to largest and smallest alleles.

et al. 1987) (table 2). In consideration of the vast number of $(CA)_n$ blocks in the human genome, it is likely that a subset of as many as several thousand can be identified with average heterozygosities of ≥70%.

Discussion

Application of the polymerase chain reaction to type polymorphic DNA markers such as the $(CA)_n$ block markers consumes less DNA and is faster than standard blotting and hybridization. Ten nanograms of genomic DNA is sufficient for routine genotyping of the $(CA)_n$ block markers (fig. 3), and the polymerase chain reaction is capable of amplifying DNA from a single template molecule (Saiki et al. 1988). The extreme sensitivity of the polymerase chain reaction means that it may soon be possible to skip the establishment of lymphoblastoid cell lines from each individual in large genetic-disease families. Enough DNA can be isolated from a single modest blood sample to type tens of thousands of $(CA)_n$ block markers.

The procedure for typing the $(CA)_n$ block markers is fast because it essentially involves only two steps, amplification and electrophoresis. Several different marker fragments can be amplified simultaneously (three markers in fig. 3), and amplified fragments from several markers can be loaded into each gel lane. Hundreds of $(CA)_n$ block markers can be typed on a single polyacrylamide gel. While this work was in progress, a report appeared (Skolnick and Wallace 1988) detailing the theoretical advantages of typing DNA polymorphisms by using techniques such as the polymerase chain reaction. The $(CA)_n$ block markers provide a means of realizing these advantages.

Variation in the number of repeats within a block of tandem repeats appears to be a universal feature of eukaryote DNA, regardless of the length of the repeat unit. Human color-vision genes have repeat lengths of 16 kb or more, repeats in malaria parasite antigen genes vary in length from 3 bp to 243 bp, and minisatellites have repeat lengths in the range of about 9–60 bp. In addition to $(CA)_n$ dinucleotide repeats, it is likely that length polymorphisms exist in other abundant simple-sequence tandem repeats, such as $(TC)_n$ and $(A)_n$ (Tautz and Renz 1984; Gross and Garrard 1986).

In summary, 10 specific human blocks of tandem $(dCDA)_n \cdot (dG\text{-}dT)_n$ repeats were tested for length polymorphism, and all 10 were found to be polymorphic. These markers were moderately to highly informative, and alleles were inherited in normal codominant fash-

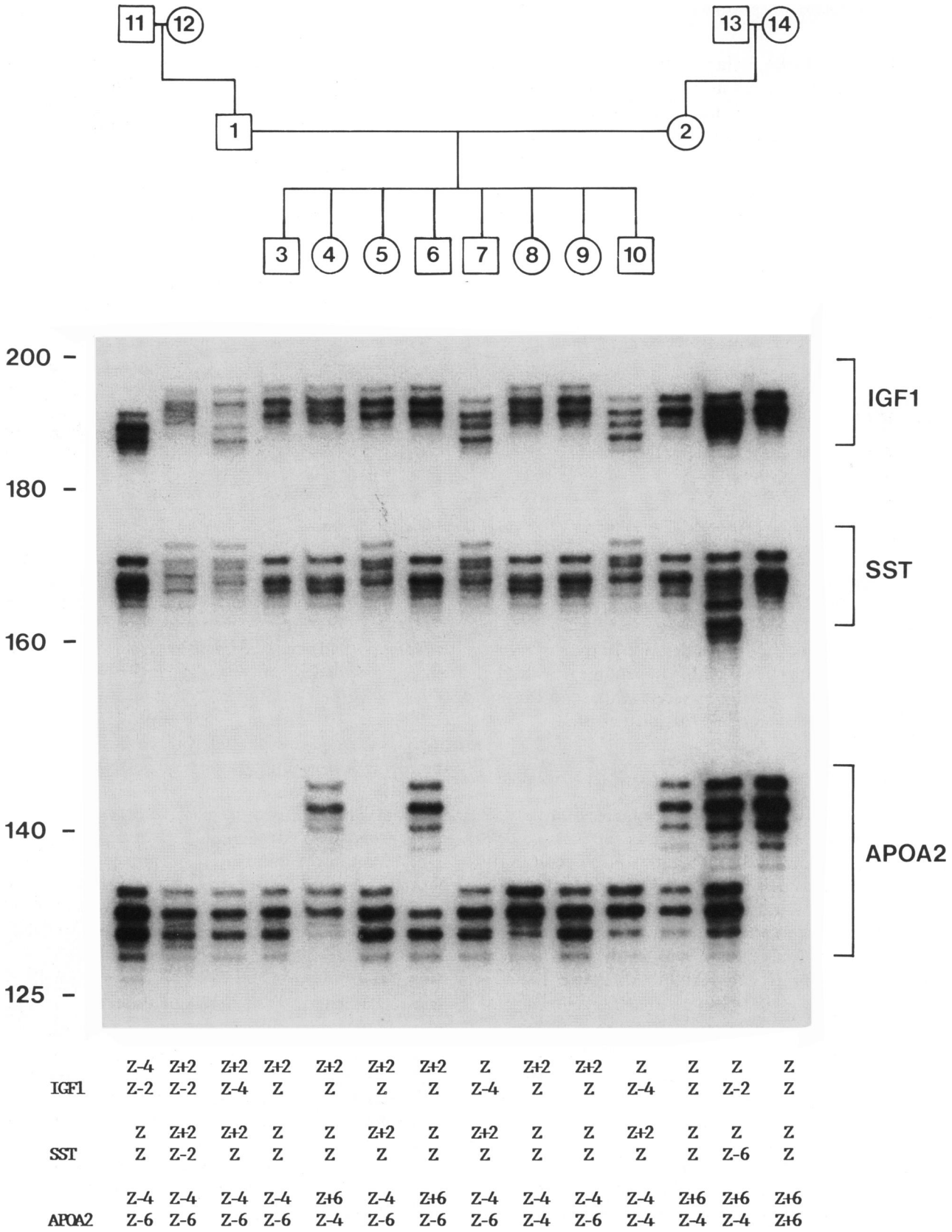

Figure 3 Inheritance of $(CA)_n$ alleles. DNA from individuals of CEPH family 1423 was amplified simultaneously with the IGF1, SST, and APOA2 primer pairs. DNA fragment sizes (in bases) are marked to the left of the gel. Individual genotypes are listed below the gel.

ion. Because of the very large number of $(CA)_n$ blocks in the human genome, this class of polymorphisms is likely to find application in the study of many genetic-disease genes and should permit substantial improvement in the resolution of the human genetic map.

Acknowledgments

We thank P. Rotwein for the control IGF1 plasmid DNA, P. de Jong and M. Van Dilla for the chromosome 19–specific phage library, J. Murray for the CEPH family DNA, J. Nicklas for DNA from the lymphocyte clones, and D. Luepke for secretarial assistance. This work was supported by the Marshfield Clinic.

References

Aldridge, J., L. Kunkel, G. Bruns, U. Tantravahi, M. Lalande, T. Brewster, E. Moreau, M. Wilson, W. Bromley, T. Roderick, and S. A. Latt. 1984. A strategy to reveal high-frequency RFLPs along the human X chromosome. Am. J. Hum. Genet. **36**:546–564.

Benton, W. D., and R. W. Davis. 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.

Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and ^{35}S label as an aid to rapid DNA sequence determination. Proc. Natl. Acad. Sci. USA **80**: 3963–3965.

Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**:314–331.

Braaten, D. C., J. R. Thomas, R. D. Little, K. R. Dickson, I. Goldberg, D. Schlessinger, A. Ciccodicola, and M. D'Urso. 1988. Locations and contexts of sequences that hybridize to Poly(dG-dT) · (dC-dA) in mammalian ribosomal DNAs and two X-linked genes. Nucleic Acids Res. **16**:865–881.

Das, H. K., C. L. Jackson, D. A. Miller, T. Leff, and J. L. Breslow. 1987. The human apolipoprotein C-II gene sequence contains a novel chromosome 19–specific minisatellite in its third intron. J. Biol. Chem. **262**:4787–4793.

Gilliam, T. C., S. T. Healey, M. E. MacDonald, G. D. Stewart, J. Wasmuth, R. E. Tanzi, J. C. Roy, and J. F. Gusella. 1987. Isolation of polymorphic DNA fragments from human chromosome 4. Nucleic Acids Res. **15**:1445–1458.

Gross, D. S., and W. T. Garrard. 1986. The ubiquitous potential Z-forming sequence of eucaryotes, $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$, is not detectable in the genomes of eubacteria. Mol. Cell. Biol. **6**:3010–3013.

Gross, D. S., S. Y. Huang, and W. T. Garrard. 1985. Chromatin structure of the potential Z-forming sequence $(dT\text{-}dG)_n \cdot (dC\text{-}dA)_n$. J. Mol. Biol. **183**:251–265.

Hamada, H., and T. Kakunaga. 1982. Potential Z-DNA forming sequences are highly dispersed in the human genome. Nature **298**:396–398.

Hamada, H., M. G. Petrino, and T. Kakunaga. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. Proc. Natl. Acad. Sci. USA **79**:6465–6469.

Hamada, H., M. G. Petrino, T. Kakunaga, M. Seidman, and B. D. Stollar. 1984*a*. Characterization of genomic Poly(dT-dG) · Poly(dC-dA) sequences; structure, organization, and conformation. Mol. Cell. Biol. **4**:2610–2621.

Hamada, H., M. Seidman, B. H. Howard, and C. M. Gorman. 1984*b*. Enhanced gene expression by the Poly(dT-dG) · Poly(dC-dA) sequence. Mol. Cell. Biol. **4**:2622–2630.

Jeffreys, A. J., N. J. Royle, V. Wilson, and Z. Wong. 1988. Spontaneous mutation rates to new length alleles at tandem-repetitive hypervariable loci in human DNA. Nature **332**: 278–281.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable 'minisatellite' regions in human DNA. Nature **314**: 67–73.

Kemp, D. J., R. L. Coppel, and R. F. Anders. 1987. Repetitive proteins and genes of malaria. Annu. Rev. Microbiol. **41**:181–208.

Miesfeld, R., M. Krystal, and N. Arnheim. 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human delta and beta globin genes. Nucleic Acids Res. **9**:5931–5947.

Motulsky, A. G. 1988. Normal and abnormal color-vision genes. Am. J. Hum. Genet. **42**:405–407.

Mullis, K. B., and F. A. Faloona. 1987. Specific synthesis of DNA *in vitro* via a polymerase-catalyzed chain reaction. Methods Enzymol. **155**:335–350.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235**: 1616–1622.

Nicklas, J. A., T. C. Hunter, L. M. Sullivan, J. K. Berman, J. P. O'Neill, and R. J. Albertini. 1987. Molecular analyses of *in vivo* hprt mutations in human T-lymphocytes. Mutagenesis **2**:341–347.

Nordheim, A., and A. Rich. 1983. The sequence $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ forms left-handed Z-DNA in negatively supercoiled plasmids. Proc. Natl. Acad. Sci. USA **80**:1821–1825.

Overhauser, J., J. McMahan, and J. J. Wasmuth. 1987. Identification of 28 DNA fragments that detect RFLPs in 13 distinct physical regions of the short arm of the chromosome 5. Nucleic Acids Res. **15**:4617–4627.

Pardue, M. L., K. Lowenhaupt, A. Rich, and A. Nordheim. 1987. $(dC\text{-}dA)_n \cdot (dG\text{-}dT)_n$ sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO J. **6**:1781–1789.

Rotwein, P., K. M. Pollock, D. K. Didier, and G. G. Krivi.

1986. Organization and sequence of the human insulin-like growth factor I gene. J. Biol. Chem. **261**:4828–4832.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.

Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim. 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science **230**:1350–1354.

Schumm, J. W., R. G. Knowlton, J. C. Braman, D. F. Barker, D. Botstein, G. Akots, V. A. Brown, T. C. Gravius, C. Helms, K. Hsiao, K. Rediker, J. G. Thurston, and H. Donis-Keller. 1988. Identification of more than 500 RFLPs by screening random genomic clones. Am. J. Hum. Genet. **42**:143–159.

Shen, L. P., and W. J. Rutter. 1984. Sequence of the human somatostatin I gene. Science **224**:168–171.

Skolnick, M. H., and R. B. Wallace. 1988. Simultaneous analysis of multiple polymorphic loci using amplified sequence polymorphisms. Genomics **2**:273–279.

Slightom, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal $^{G}\gamma$- and $^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell **21**:627–638.

Sun, L., K. E. Paulson, C. W. Schmid, L. Kadyk, and L. Leinwand. 1984. Non-Alu family interspersed repeats in human DNA and their transcriptional activity. Nucleic Acids Res. **12**:2669–2690.

Tautz, D., and M. Renz. 1984. Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. **12**:4127–4138.

Tyler-Smith, C., and W. R. A. Brown. 1987. Structure of the major block of alphoid satellite DNA on the human Y chromosome. J. Mol. Biol. **195**:457–470.

Waye, J. S., S. J. Durfy, D. Pinkel, S. Kenwrick, M. Patterson, K. E. Davies, and H. F. Willard. 1987. Chromosome-specific alpha satellite DNA from human chromosome 1: hierarchial structure and genomic organization of a polymorphic domain spanning several hundred kilobase pairs of centromeric DNA. Genomics **1**:43–51.

Weber, J. L. 1988. Molecular biology of malaria parasites. Exp. Parasitol. **66**:143–170.

Weber, J. L., J. A. Lyon, R. H. Wolff, T. Hall, G. H. Lowell, and J. D. Chulay. 1988. Primary structure of a *Plasmodium falciparum* malaria antigen located at the merozoite surface and within the parasitophorous vacuole. J. Biol. Chem. **263**:11421–11425.